

BoVWを用いた室内における空間特徴量抽出法の提案

○矢口瑛貴[†] 久留亜沙美^{††} 楨原絵里奈[†] 小野景子[†]
([†]同志社大学理工学部, ^{††}同志社大学理工学部研究科)

Space Ambiance Extraction using Bag of Visual Words

*A. Yaguchi[†], A. Kuru^{††}, E. Makihara[†] and K. Ono[†]

([†] Faculty of Science and Engineering, Doshisha University

^{††} Graduate School of Science and Engineering, Doshisha University)

Abstract— In text-based online searches for furniture, it is difficult to estimate a space ambiance, which affects the combinations with other furniture. Various methods using shape features and textures have been proposed to estimate the space ambiance, however these methods need to estimate the optimal number of features. This study estimates the space ambiance by extracting visual words from in-room image data, considering each image as a single sentence. Specifically, we use the Bag of Visual Words, a natural language processing method, to estimate the space ambiance using a histogram based on the extracted visual words.

Key Words: Bag of Visual Words, Feature extraction, Furniture, Histogram, SURF

1 はじめに

テキスト入力に基づいたオンライン検索では、画像や動画など様々な情報を検索、推薦可能である¹⁾。しかし、物品毎の関連性や好みに影響がある家具や洋服は、ユーザの好みを反映して検索、推薦することは容易でない²⁾。そのため、現行のテキスト検索ではなく、ユーザの好みを反映可能なオンライン検索システムの需要が増加している³⁾。その解法の一つとして、画像特徴量に基づく検索システムが提案されており、画像に基づく好みや類似性からユーザの好みを反映した推薦が可能である⁴⁾。

本研究では、家具に着目して研究を行う。従来より、家具に着目した推薦システムがあるが、従来のシステムでは、家具を推薦する際に、家具や植物、壁、床などを含めた部屋空間の雰囲気（以後、空間の雰囲気）を考慮していないため、ユーザの好む家具を推薦する際に問題が生じていた⁵⁾。空間の雰囲気は、ユーザの好みに重要な要素であると報告されており、ユーザの好む家具を推薦するためには、空間の雰囲気を推定する必要があると考えられる⁴⁾⁶⁾⁷⁾。空間の雰囲気を推定するため、空間を構成する物体の形状特徴やテクスチャなどを用いた手法が提案されている。しかし、空間の雰囲気は空間と家具、家具同士の関係や家具のテクスチャなど考慮する要素が多いため、最適な空間の特徴量を推定することは容易でない⁶⁾。

そこで、本研究では、新たな空間特徴量抽出法を提案する。具体的には、画像から空間特徴量を抽出し、空間の雰囲気の推定において考慮する要素を削減する。画像から雰囲気を推定する手法として、Bag of Visual Words（以後、BoVW）を用いた手法が有効であるか検証する。

2 先行研究

2.1 オノマトペと質感に基づく家具推薦システム

画像から抽出した質感および色特徴量を質感に関する言語（以後、オノマトペ）と対応させる手法が提案されている。システムの概要図を Fig. 1 に示す。この手法は、オノマトペと質感特徴量、オノマトペと色特徴

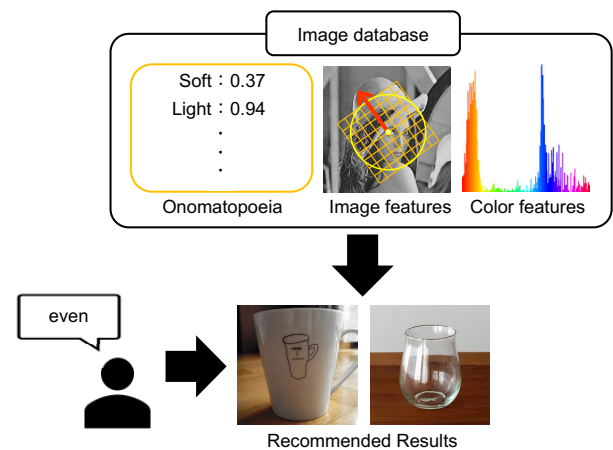


Fig. 1: Overview of onomatopoeia and texture-based recommendation system

量との類似度を算出することにより、オノマトペに適した画像のランク付けを行う。オノマトペのようなテキストデータだけでなく、画像特徴量、色特徴量を用いることでユーザ好みの商品を推薦を可能にした⁵⁾⁸⁾。

2.2 リビングと家具の形容詞間の相関に基づく家具推薦システム

家具を含まないリビング画像と三次元椅子データの形容詞を用い、リビングと椅子の関係を表現することで、推薦システムを構築している⁹⁾。画像と形容詞の対応付けは、人の評価実験によって行う。さらに、正準相関分析を用いることでリビングに調和した椅子の推薦を可能にした。正準相関分析とは、計測方法の異なるデータから共通成分を抽出する解析手法である。

2.3 画像特徴量抽出に関する研究

画像検索システムにおける画像特徴量抽出手法として、BoVWを用いた手法が有効であると報告されている¹⁰⁾。BoVWの概要図を Fig. 2 に示す。BoVWとは、自然言語処理分野における文章の特徴量抽出法として一般的であるBag of Wordsを画像に応用した手法である。Bag of Wordsとは、TF-IDFで特徴語を抽出し

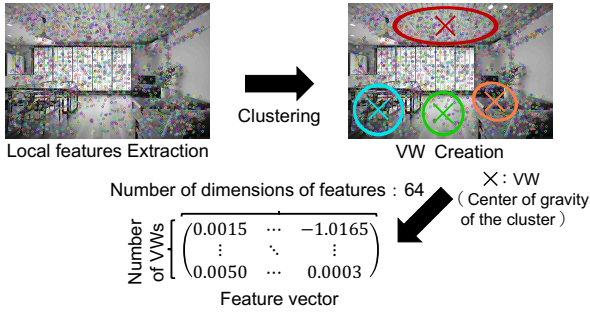


Fig. 2: BoVW

て文書ベクトルの次元とし、TF-IDF の重みを文書ベクトルの値とする手法である¹¹⁾。BoVW では、Bag of Words の文書が画像、単語が局所特徴量に対応する。具体的には、画像から局所特徴量を抽出し、局所特徴量を K 個にクラスタリングする。そして、作成した K 個の各クラスターの重心を Visual Word (以後、VW) とし、画像を特徴ベクトルである VW の集合とみなす。画像のカテゴリ毎に VW を作成することで、カテゴリの追加に伴う検索率低下を是正可能であると報告されている¹⁰⁾。

BoVW では、局所特徴量として、Scale-Invariant Feature Transform (以後、SIFT) 特徴量や SIFT 特徴量の高速版である Speeded-Up Robust Features (以後、SURF) 特徴量を用いる。局所特徴量とは、画像中の輝度勾配の変化が大きい特徴点を検出し、検出した特徴点周りの領域を画素値や微分値に基づき導出した特徴ベクトルである⁸⁾。SIFT 特徴量抽出の概要図を Fig. 3 に、Difference of Gaussian (以後、Dog) を用いた特徴点検知を Fig. 4 に示す。SIFT 特徴量抽出は、Fig. 4 に示す画像の差分画像である Difference-of-Gaussian により輝度勾配の変化を算出し、特徴点検出を行う。SURF 特徴量は、特徴点検出に用いる Hessian-Laplace 検出器を Box フィルタで近似することで高速化する。なお、Box フィルタとは、注目画素を中心とした正方形の範囲の画素の平均を算出するフィルタである。また、Hessian-Laplace 検出器にはヘッセ行列を用いる。ヘッセ行列を以下の式 (1) に示す。

$$H(x, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{xy}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix} \quad (1)$$

ここで、 $L_{xx}(x, \sigma)$ は画像中の点 x におけるガウシアン二次微分 $\sigma^2/\sigma x^2 G(\sigma)$ の畳み込み結果である。 $L_{xy}(x, \sigma), L_{yy}(x, \sigma)$ についても同様である。ヘッセ行列によって、近似度合いが変化するため、SURF 特徴量のパラメータであるヘッセ閾値が存在する。ヘッセ閾値の大小により、抽出する SURF 特徴量数の調整が可能である。

3 提案手法

提案手法は、画像を局所特徴量のヒストグラムと捉え、BoVW およびヒストグラムを用いて空間全体の特徴量を抽出する。提案手法の概要図を Fig. 5 に示す。性能評価には、cos 類似度、Accuracy、Recall、K-分割交差検証法を用いる。具体的には、BoVW よって VW を作成後、局所特徴量を近似する VW へ振り分けることで VW 数を次元とするヒストグラムを作成する。そ

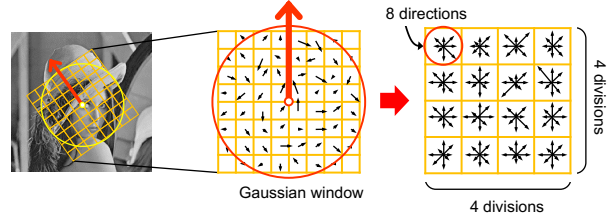


Fig. 3: Local feature extraction

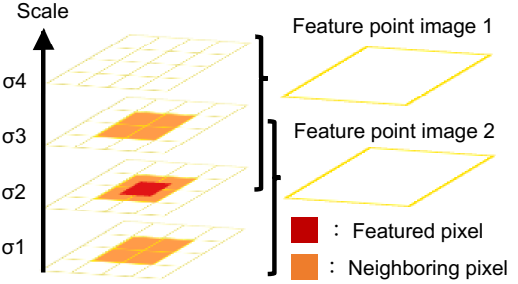


Fig. 4: Detection of feature points using Dog

して、ヒストグラム間 cos 類似度を算出し、性能評価を行う。提案手法の手順を以下に示す。

1. I カテゴリに各 J 枚存在する学習画像 $S = \{S_{ij}; i = 1, \dots, I, j = 1, \dots, J\}$ の各 S_{ij} に対する SURF 特徴量 $F_{ij} = \{f_{ij}^k; k = 1, \dots, K\}$ を抽出する。
2. 全学習画像の SURF 特徴量 \mathcal{F} をクラスタリングし、学習画像の VW である $W = \{w_l; l = 1, \dots, L\}$ を導出する。
3. SURF 特徴量 F_{ij} から W に基づく各学習画像のヒストグラム $Hs_{ij} = \sum_L w_l$ を作成する。
4. Hs_{ij} から各カテゴリの平均ヒストグラム $H_i = \frac{1}{J} \sum_J Hs_{ij}$ を作成し、正解ヒストグラム $H = \{H_i\}$ を作成する。
5. I カテゴリに各 N 枚存在する $M = I * N$ 枚のテスト画像 $T = \{T_m; m = 1, \dots, M\}$ から W に基づく各テスト画像のヒストグラム HT_m を作成する。そして、テスト画像のヒストグラム $HT = \{HT_m\}$ を作成する。
6. H と HT から cos 類似度 $\cos_m^i = Sim(H, HT_m)$ を算出し、分類結果 $\hat{I} = \arg \max_i \cos_m^i$ を導出する。なお、 $Sim(x, y)$ は x, y の cos 類似度を算出する関数である。その後、 \hat{I} に基づき T を分類する。
7. T の分類結果に基づき、性能評価を行う。

4 提案手法の性能評価検証

4.1 検証目的

本検証では、BoVW を用いた室内における空間特徴量抽出法が有効か検証するため、従来手法である Naive 手法と比較して、空間特徴量抽出における提案手法の有効性を検証する⁵⁾。また、空間特徴量を大域特徴量として定量化可能であるか検証する。

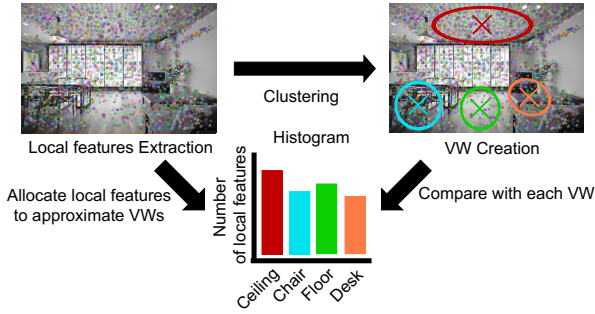
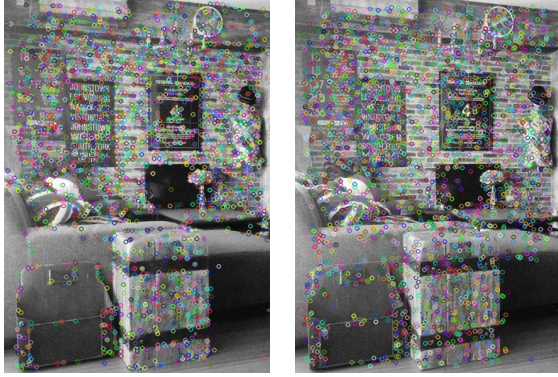


Fig. 5: Overview of the proposed method



(a) SIFT features (b) SURF features
Fig. 6: Comparison with SIFT and SURF features

4.2 検証方法

Naive手法と提案手法の Accuracy を比較することにより、空間特徴量抽出における提案手法の有効性を検証する。なお、空間特徴量とは、壁のテクスチャや床のテクスチャ、家具や植物、ランプなどのアイテムを指す。使用画像に対する SIFT 特徴量および SURF 特徴量を Fig. 6 に示す。使用する画像に対して SIFT 特徴量および SURF 特徴量の比較を行った結果、Fig. 6 のように特徴量数に大きな差異がない。そのため、局所特徴量には SIFT 特徴量の高速版である SURF 特徴量を用いる。なお、クラスタリングには計算コストを抑えるため、非階層クラスタリング手法である K-means 法を用いる。また、性能評価には cos 類似度と K-分割交差検証法を用いる¹²⁾。

4.3 比較手法

提案手法と比較する Naive 手法は、画像を VW の集合と捉え、BoVW を用いて空間全体の特徴量を抽出する手法である。性能評価には、cos 類似度、Accuracy、Recall、K-分割交差検証法を用いる。具体的には、BoVW によって作成した VW の VW 間 cos 類似度を算出し、性能評価を行う。Naive 手法の手順を以下に示す。

1. I カテゴリに各 J 枚存在する学習画像 $S = \{S_{ij}; i = 1, \dots, I, j = 1, \dots, J\}$ の各 S_{ij} に対する SURF 特徴量 $F_{ij} = \{f_{ij}^k; k = 1, \dots, K\}$ を抽出する。
2. I カテゴリの画像の SURF 特徴量 F_i をクラスタリングし、学習画像の VW である $W_i = \{w_i^l; l = 1, \dots, L\}$ を導出する。
3. I カテゴリの各 N 枚の $M = I * N$ 枚のテスト画像 $T = \{T_m; m = 1, \dots, M\}$ の各 T_m に対する VW

である $WT_m = \{wt_m^l; l = 1, \dots, L\}$ を導出する。

4. W_i と WT_m の VW の対応付け $\hat{l} = \arg \max_l Sim(W_i^l, WT_m^l)$ を導出する。ここで、 $l' = 1, \dots, L$ 、 Sim は x, y の cos 類似度を算出する関数である。 l' に基づき、 W_i を並べ替えた \hat{W}_i を作成する。
5. \hat{W}_i と WT_m から cos 類似度 $\cos_m^i = Sim(\hat{W}_i, WT_m)$ を算出し、分類結果 $\hat{I} = \arg \max_i \cos_m^i$ を導出する。なお、 $Sim(x, y)$ は x, y の cos 類似度を算出する関数である。その後、 \hat{I} に基づき T を分類する。
6. T の分類結果に基づき、性能評価を行う。

4.4 使用する画像データセット

使用する画像データセットを Fig. 7 に示す。本検証で使用する画像データセットは、Fig. 7 に示すように、Brick tone, Japanese modern tone, Wood tone の部屋画像 3 カテゴリに分類し、Bing 画像検索 API より自動収集する。Bing 画像検索 API のキーワードに依存する収集精度の差を是正するため、Bing 画像検索 API によって収集した画像と合成画像で構成する。合成画像を Fig. 8 に示す。Fig. 8 に示すように、合成画像は、各カテゴリの画像に対して、椅子や机、植物など不足するアイテムを追加し、作成した。合成画像の作成方法を Fig. 9 に示す。合成画像を作成するための条件は、部屋空間に椅子、机、植物が存在する部屋とする。

本研究では、ユーザの好みの要素の一つである空間の雰囲気を出すため、各カテゴリ 50 枚 (計 150 枚) を主観で選別し、使用する。なお、40 枚は学習用画像、10 枚はテスト用画像とする。

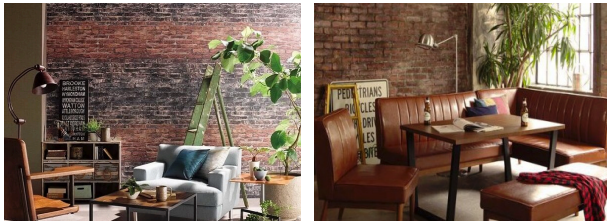
4.5 クラスタ数

局所特徴量の分類数であるクラスタ数の範囲は 2~8 とし、クラスタ数を 1 ずつ変化させて検証を行う。クラスタ数 2 とは、クラスタリングにおける分類数の最小値である。また、クラスタ数 8 とは、壁のテクスチャ、床のテクスチャ、家具、アイテムなどを想定する分類対象の最大値である。

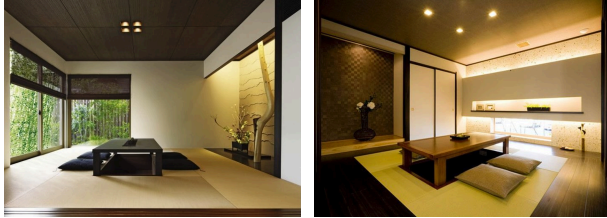
4.6 検証結果および考察

Naive 手法および提案手法の Accuracy を Fig. 10 に示す。Fig. 10 より、Naive 手法と比較して提案手法は、クラスタ数 3~8 の場合、Accuracy が高かった。Naive 手法では、空間特徴量を画像の VW から推定するため、画像に対する VW の重要度付けが困難である。一方で、提案手法では、空間特徴量を画像のヒストグラム情報から推定するため、画像に対する VW の重要度付けが容易である。そのため、Naive 手法と比較して提案手法の Accuracy が高かったと考えられる。

クラスタ数 2~8 の場合の提案手法の Recall を Fig. 11 に示す。Fig. 11 より、提案手法の Accuracy が最も高いクラスタ数 7 の場合であっても、Brick tone の Recall は 0.96 と高いが、Wood tone および Japanese modern tone の Recall は 0.7 未満と低い。Japanese modern tone の画像を Fig. 12 に示す。Fig. 12 より、Japanese modern tone の画像には、フローリングが含まれる。そのため、Wood tone の画像と共通のフローリングとい



(a) Brick tone

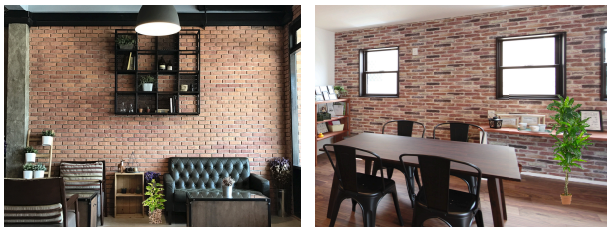


(b) Japanese modern tone

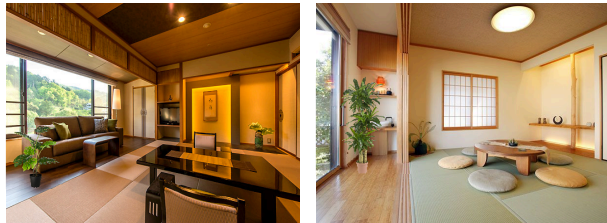


(c) Wood tone

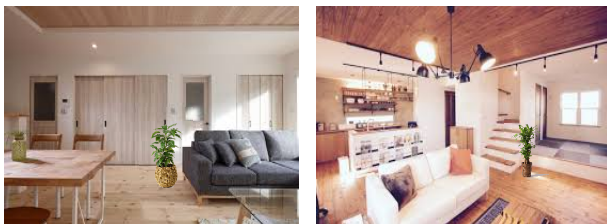
Fig. 7: Image data set



(a) Brick tone



(b) Japanese modern tone



(c) Wood tone

Fig. 8: Artificial images

う特徴が存在し、Wood tone および Japanese modern tone の分類が困難であり、Recall が低下したと考えられる。また、本検証で設定したクラスタ数 2~8 には有効なクラスタ数が存在せず、Wood tone および Japanese modern tone の Recall が低かったと考えられる。

以上より、提案手法は、空間の雰囲気画像のヒスト

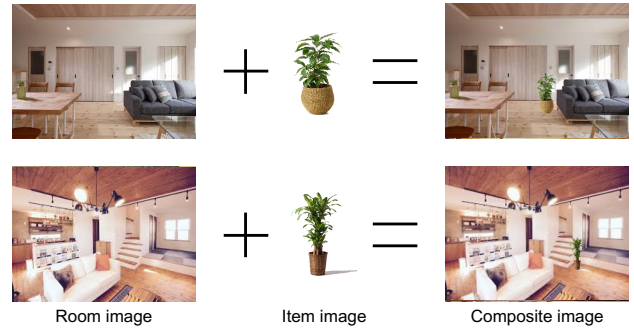


Fig. 9: How to create artificial images

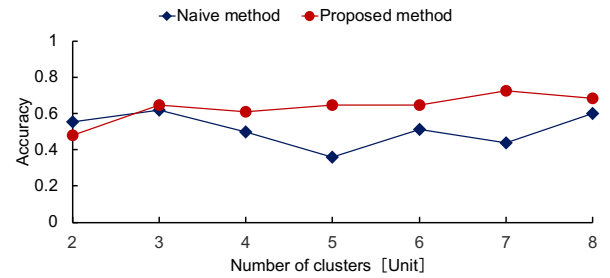


Fig. 10: Comparison with Accuracy of Naive and Proposed methods

グラム情報から推定するため、家具やアイテムなどにとられない特徴量抽出が可能であることがわかった。そのため、提案手法は空間特徴量抽出法として有効であり、空間特徴量を大域特徴量として定量化可能であると考えられる。また、クラスタ数 2~8 では、Brick tone の特徴量抽出は可能だが、Japanese modern tone および Wood tone の特徴量抽出は容易でないことがわかった。

5 提案手法における有効なクラスタ数の分析

5.1 分析目的

提案手法が BoVW に基づく手法であるため、クラスタ数の増加に伴い、画像の特徴量の分類数は増加する。そのため、クラスタ数によって、空間特徴量の抽出度が変動すると考えられる。そこで、提案手法における空間特徴量抽出に有効なクラスタ数を検証する。

5.2 分析方法

提案手法のクラスタ数に伴う Accuracy と Recall を比較することにより、提案手法における有効なクラスタ数を分析する。クラスタリング手法および使用する画像データセットは、性能評価検証と同様である。

5.3 クラスタ数

クラスタ数の範囲は 2~294 とし、クラスタ数を 1 ずつ変化させて分析を行う。クラスタ数 2 とは、クラス

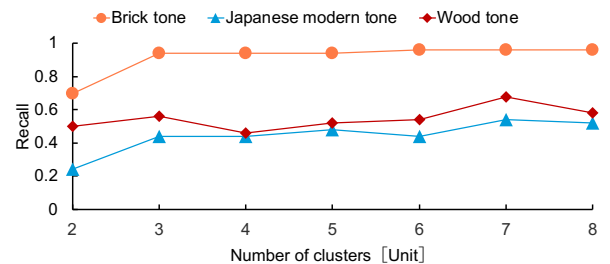


Fig. 11: Recall for the proposed method



Fig. 12: Japanese modern tone images

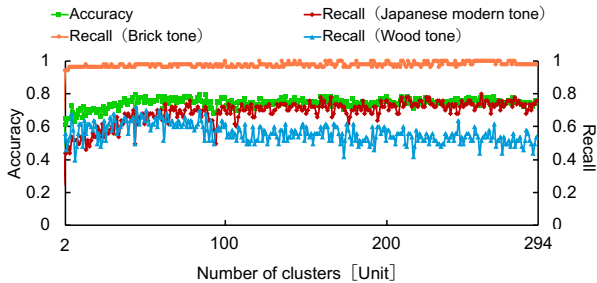


Fig. 13: Accuracy and Recall for the proposed method with the number of clusters from 2 to 294

タリングにおける分類数の最小値である。また、クラスタ数 294 とは、使用する画像から得られる SURF 特徴量数の最小値である。

5.4 分析結果および考察

提案手法におけるクラスタ数 2~294 の Accuracy および Recall を Fig. 13 に示す。Fig. 13 より、クラスタ数 47 の時に、Accuracy が 0.79 と最も高くなった。また、クラスタ数が 4 から 47 の場合は、クラスタ数が増加するにつれて、全てのカテゴリにおいて Recall が増加傾向にある。一方、クラスタ数が 48~294 の場合は、Brick tone と Japanese modern tone の Recall は上昇傾向であるが、Wood tone のみ Recall が減少傾向となった。各カテゴリの SURF 特徴量数を Table 1 に示す。Wood tone の Recall が 48~294 の場合で減少傾向にある理由として、Table 1 より、使用した画像データセットの Brick tone, Japanese modern tone と比較して、SURF 特徴量が少なかったためであると考えられる。

以上より、使用した画像データセットに対して、提案手法における有効なクラスタ数は 47 であり、クラスタ数の調整により、性能の向上が可能であることがわかった。

6 おわりに

本研究では、空間特徴量を大域特徴量として抽出する提案手法の有効性および提案手法における有効なクラスタ数を検証した。検証の結果、Naive 手法と比較して、提案手法は Accuracy の性能が良かったため、空間特徴量抽出法として有効であることがわかった。また、使用した画像データセットに対して、クラスタ数を調整することで、性能の向上が可能であるとわかった。

以上より、室内における空間特徴量を画像から抽出した局所特徴量のヒストグラム情報から推定する手法は有効であることを示した。また、空間特徴量は大域特徴量として定量化可能と確認した。そして、空間特徴量を考慮することにより、ユーザ趣向に基づく家具の推薦が可能になると考えられる。

Table 1: Number of SURF features in each category

	SURF features [Unit]
Brick tone	350,566
Japanese modern tone	264,470
Wood tone	229,478

参考文献

- 1) 小野寺翼：ソーシャルメディア時代の広報、情報の科学と技術, Vol.68, No.4, pp.156-159, (2018) .
- 2) 小林亜樹：抽象的なイメージによる画像検索, 映像情報メディア学会誌, Vol.59, No.8, pp.1156-1161, (2011) .
- 3) 小柴等, 相原健郎ほか：説得性に基づく情報推薦手法の提案：送り手の属性に着目したモデルと検証, 情報処理学会論文誌, Vol.51, No.8 (2010) .
- 4) 中林拓馬, 神山一馬, 黒澤紀之, 平沢岳人：画像をクエリとした建築情報検索システムの実現に向けた類似画像検索技術の検証, 日本建築学会技術報告集, Vol.19, No.42, pp.775-778, (2013) .
- 5) 土斐崎龍一, 飯場咲紀ほか：オノマトベと質感印象の結び付きに着目した商品検索への画像・テキスト情報活用の可能性, 人工知能学会論文誌, Vol.30, No.1, pp.124-137, (2015) .
- 6) 佐藤昌子, 中田美奈子, 松崎雅子：壁材と室内空間イメージ. 大阪市立大学生活科学部紀要, Vol.43, No.6, pp.33-48, (1995) .
- 7) 柴田滝也：街路景観の心理的効果のモデル化と画像検索システムへの応用, 日本建築学会計画系論文集, Vol.64, No.515, pp.297-304, (2017) .
- 8) 藤吉弘亘, 安倍満：局所勾配特徴抽出技術, 精密工学会誌, Vol.77, No.12, pp.1109-1116, (2012) .
- 9) 山際央明, 柴田滝也：データベースを用いたインテリア・コーディネート・エージェントに関する研究, 感性工学研究論文集, Vol.7, No.4, pp.685-692, (2010) .
- 10) 田中雄翔, 滝口哲也, 有木康雄：ウェブ画像を用いたカテゴリ別 Visual Words による未知物体判別, 画像の認識・理解シンポジウム (MIRU) (2012) .
- 11) Yun-tao,Zhang et al : An improved TF-IDF approach for text classification,Journal of Zhejiang University-Science, Vol.6, No.1, pp.49-55, (2005) .
- 12) 難波英嗣：テキスト間の類似度の測定, 情報の科学と技術, Vol.70, No.7, pp.373-375, (2020) .